Beyond the 'Bayesian blur': predictive processing and the nature of subjective experience

Article (Accepted Version)

Clark, Andy (2018) Beyond the 'Bayesian blur': predictive processing and the nature of subjective experience. Journal of Consciousness Studies, 25 (3-4). pp. 71-87. ISSN 1355-8250

Beyond the 'Bayesian Blur': Predictive Processing and the Nature of Subjective Experience*

Andy Clark[1,2]

Abstract

Recent work in cognitive and computational neuroscience depicts the brain as in some (perhaps merely approximate) sense implementing probabilistic inference. This suggests a puzzle. If the processing that enables perceptual experience involves representing or approximating probability distributions, why does experience itself appear univocal and determinate, apparently bearing no traces of those probabilistic roots? In this paper, I canvass a range of responses, including the denial of univocality and determinacy itself. I argue that there is reason to think that it is our conception of perception itself that is flawed. Once we see perception aright, as the slave of action, the puzzlement recedes. Perceptual determinacy reflects only the mundane fact that we are embodied, active, agents who must constantly engage the world they perceptually encounter.

[1] School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, UK
[2] Dept of Philosophy, Macquarie University, Sydney, Australia

1. Predictive Brains

Brains like ours, according to some highly influential current theorizing in cognitive and computational neuroscience, are multi-level prediction machines – 'predictive processors' (Clark (2016)) that approximate Bayesian forms of inference as they deploy stored knowledge to negotiate a complex and uncertain world[1]. Bayesian inference is an optimal means of making use of new information in ways that are properly sensitive to uncertainty (for a nice introduction, see Wiese (2016)). Predictive processing approximates Bayesian methods by combining prior knowledge with incoming sensory evidence in ways that take account of uncertainty. Such systems are constantly trying to guess the incoming sensory barrage at many scales of space and time. To see a complex structured world of interacting objects is here to deploy a predictive model (a so-called 'generative model') able to construct the incoming sensory signal for itself, using stored knowledge about the world.

Momentary perceptual experience, if these stories are on track, always reflects a delicate combination of top-down model-based prediction, self estimated sensory uncertainty, and bottom-up (incoming) sensory evidence. When the top-down prediction is wrong, 'prediction error' signals result, and these are fed upwards (and sideways), allowing new 'top-down' guesses rapidly to be recruited. It is only when all that settles, within current tolerances of noise, that a clear percept is formed. Given some sensory input ('evidence'), such systems seek out the set of interacting worldly causes that make that evidence most likely. In essence, this is to attempt to guess the incoming sensory signal itself, given probabilistic knowledge about how the world is most likely to be. This process of 'guessing the signal' takes place at every level of neural processing, with each higher level constantly trying to guess the activity of the level immediately below (which it is thus treating as if it were a raw sensory signal).

In pure Bayesian inference[2], a system would not compute a single value, but rather deliver a conditional probability density function. Thus suppose the goal is to discover the depth of a visible object. Instead of computing a single value, such as '6 inches distant', the conditional probability density function would encode "the relative probability that the object is at different depths Z, given the available sensory information" (Knill and Pouget 2004: 712). For example, it might suggest that the object is 6 inches distant with 70% probability, 7 inches distant with 20% probability, and 5 inches distant with 10% probability.

It is extremely unlikely, however, that real brains compute such functions. Instead, they must rely on approximations whose form has yet to be conclusively established. Possible approximations include ones that represent only the mean and the variance of the distribution (it's 'sufficient statistics' – see e.g Wiese (2016)). The lack of an agreed answer to the question of what approximations the brain actually uses – assuming it is indeed approximating Bayesian inference - has been dubbed (Wiese (2016)) the Probability Conundrum. However, even if our predictive processing brains are not explicitly representing probability values or densities, they are still combining prior knowledge with new evidence in ways that respect (by approximating) probabilistic Bayesian inference. The situation is nicely captured by John Kruschke (a leading Bayesian cognitive scientist) who, after describing standard approaches as associating descriptors with single determinate values, writes that:

> "Bayesian approaches assume a radically different mental ontology, in which the learner entertains an entire spectrum of hypothetical values for every descriptor. For example, the association between "shock" and "tone" might be anything on an infinite continuum, and the learner's knowledge consists of a distribution of believabilities over that continuum. The learner may believe most strongly in a value of 0.413 but also have some belief in values larger or smaller. Entertaining an infinite space of hypothetical values does not imply the need for an information processor of infinite capacity, for infinite belief distributions can be represented with small sets of values. For example, a normal distribution, which extends over an infinite space, is fully represented by its mean and variance." Kruschke (2008) p.210-211

This leads to a potential puzzle, as we shall next see.

2. The Puzzle of Unitary-Coherent Perceptual Experience

In the process of perceiving the world we seem to experience, moment-by-moment, a clear and univocal 'way things are', rather than what Lu et al (2016) colourfully describe as a 'Bayesian blur' of possibilities'[3]. More precisely, experience seems to be 'unitary-coherent' (UC) where:

> Unitary means we only perceive one interpretation at a time (e.g. either a face or a vase in the Rubin Vase illusion) rather than a blur of multiple possible interpretations (never the face and vase together). Coherent means that we almost always perceive scenes comprised of parts which do not contradict one another (e.g we do not see part face and part vase). Lu et al (2016) p.259

It might be thought that the use of certain approximations dissolves the puzzle[4]. But although the underlying representations might now be of (for example) simply the mean and variance of the distribution, they still capture – as Kruschke insists - the broad shape of a probabilistically inflected space. Indeed, if they failed to do so, they would be unable even to approximate probabilistic Bayesian inference. So although there is room to ask what difference different approximations might make, the core puzzle (I argue) is more general, and arises for every story that might reasonably be counted as an approximation to probabilistic inference. Moreover, even though it is true that in predictive coding formulations beliefs over continuous states of affairs are assumed to be Gaussian and can therefore only have one peak, PP itself depicts a process of hierarchical Bayesian inference in which there can be multiple models, each with their own peak. The core puzzle, carefully expressed, thus concerns why moment-to-moment experience is unitary-coherent and seems to bear no traces of all this multi-level probabilistic guessing?

Notice that this is not even simply a question about why there is a single winning interpretation. For there could be a single winning interpretation yet some form of experience that still displays traces of the space of near-misses – faint images, perhaps, of nearby ways the world might have been. Such a possibility seems especially pertinent when we consider (section 4) cases of bistable perception. Moreover, it is widely held that it is the addition of a utility or cost function (a constraint on acceptable solutions) that forces the posterior density function to a single value[5]. In motor control, for example, jerkiness is a cost that is minimized by good, smooth, solutions, and avoiding undue jerkiness can be used to help select a single solution from the space of possibilities. So even if perceptual inference involves somehow computing over multiple probabilities the perceptual 'output', once constrained by a cost function, need not reflect this fact.

It is worth noting, however, that things look somewhat different from the perspective of the PP theorist. This is because PP (in the version defended by Friston, Hohwy, Clark, Seth, and others) eschews the appeal to cost and utility functions entirely. The reasons for this are complex (for discussion, see Friston (2011), Friston et al (2012), Clark (2016), chapter 4). But in essence, the idea is to absorb such functions into the generative model itself – a model that then simply predicts that motor behaviors will minimize 'costs' such as jerkiness or rate of change of torque. This sidesteps the need for certain expensive computations (see Friston (2011) pp. 491-492). Nonetheless, the effect of incorporating the right expectations (in this case, expectations of smooth reaching etc.) into the generative model is the same, since prior beliefs about the (e.g. non-jerky) shape of trajectories of motion can now determine unique results.

The puzzle is thus not how PP-style probabilistic processing could yield determinate, unitary-coherent results in the case of general perception, but *why it should do so*. In bringing about smooth, world-engaging motion, the value of avoiding jerk (and so building jerk-minimizing predictions into the generative model) is clear. But why, when experiencing the world, is perceptual processing being forced to depict only a single determinate and unified way things are? Why not keep alive, in experience, weighted traces of all the possibilities that are supported by the sensory evidence? Surely not to do so is to suppress important information that might usefully inform reason and choice?

At this point it might be suggested that experience does, at least at times, bear such traces. I return to this issue in Section 3. For present, I assume it does not. The puzzle I want to foreground can now be properly expressed. It is this. Why, assuming that perceptual experience is rooted in computations that trade (explicitly or implicitly) in probabilities, *and* only the PP-apparatus of generative model-based prediction, does experience turn out to be unitary-coherent? Why, in other words, is the posterior probability distribution apparently forced to collapse to a determinate (unitary-coherent) 'take' on how things are?

The same puzzle is informally expressed by Holton who notes that:

> "At the level of what we see, rather than that of what our unconscious visual systems are doing, we don't have a graded continuum of confidence in different hypotheses. Perceptions are all-or-nothing." Holton 2016: 10

These puzzles about perception have a possible parallel in the case of intentions. Thus Holton (op cit) suggests that here too "at the level of what subjects themselves do… the relevant state looks to be typically all-or-nothing: either I intend to do something or I don't." The worry here is that predictive processing models treat intentions to act as *high-level probabilistic predictions*, some of which win out over others so as to bring about actions[6]. But when I intend to pick up the coffee and drink it, it does not feel as if I predict the act (or anything else, really) with only, say, 80% probability. Here too, it looks as if the agentive experience has been forced into a kind of determinacy and univocality that the bedrock processing lacks.

A basic puzzle has now emerged. If the computations underlying perception and action are probabilistic (either involving explicit representations of probabilities or handy approximations), why does human experience present a univocal (unitary-coherent) scene, and why do our intentions to act strike us as determinate and univocal rather than probabilistic? It can seem (Holton (2016), Klein (In Press)) as if the sub-personal story here is somehow failing to make sufficient contact with the so-called 'personal-level' (Dennett (1969)) facts – facts about how the world, including our own mental life, appears to us as a reasoning and thinking being

3. Perception in Action

A deflationary response would be to argue that experience and intentions are not so unitary, univocal, or determinate after all – that they do bear traces of their putatively probabilistic roots. In this vein, Morrison (2016) (2017) argues that degrees of confidence are 'assigned' within perceptual experience itself. In other words, degrees of confidence are not merely the results of a doxastic or post-perceptual (e.g. meta-cognitive) exercise but are routinely given within perceptual experience.

Even if this is the case, however, our core puzzle remains. For however such confidence may be manifest (assuming it is) in experience, it is surely not manifest as e.g. concurrent *visual* experience of other (nearby) ways the current scene might be. A scene may look fuzzy or unclear, and I may have more or less confidence in who or what it is that I am seeing. But that fuzziness and lack of confidence is itself given via or within a single unitary-coherent percept. There is still, if you like, just one way the scene currently looks to me – a way that is unitary-coherent even on those occasions when the scene seems fuzzy or unclear.

There are other kind of case we might consider too. It might be claimed that object distance (for example) doesn't look determinate at all, because we may well be uncertain between multiple estimates of that distance. But while this may be true for estimations as expressed (say) in inches, that does not yet imply that the object itself visually appears to be at an uncertain depth. The uncertainty here may simply be about how to translate a determinate impression of depth into a public measurement system. There may be other times, however, when we do feel a genuine sense of uncertainty about the nature of the experience or percept itself, and not just how to label or categorize it. For example, I may be unsure just how I feel about going on a trip, or whether or not I can detect some very faint pattern. These are interesting and revealing cases, and I return to them later on.

Our perceptual worlds display unity and coherence, and depict a single way things are. The reason for this lies, I now want to suggest, in the transformative role of action itself. It is the need for perception constantly to mandate action and choice that requires the system to opt for a single 'overall take' on how the world is now most likely to be. Thus, in the mostly unlikely event of a genuine tie between two or more posterior 'takes' on how things are, such a system would in effect have to toss a mental coin and go for one take rather than any other.

At first sight, this seems like a pointless refusal to profit from good information. Suppose you must decide between several possible items to purchase, and you had made an exhaustive list of all the relevant criteria. If you then provided full information about each item and asked an artificial neural network to crunch the numbers, you would surely want to be informed of the fact that three items (say) had tied for top position. If the network just tossed a coin between those three and suggested only one to you, you would be under-informed. Likewise if it 'solved' the choice puzzle using some simple but irrelevant heuristic like 'choose the one whose name has the least number of letters'.

But consider what happens next. Now, armed with the information that the three items are tied for first place, you still have to decide what to do. In many biologically realistic situations, this will be a matter of great urgency. Given (by hypothesis) that the best you can is to toss a coin, you still have to opt for one action or another!

Perhaps, then, evolution (or life-time learning) has simply arranged to by-pass that potentially time-wasting space for further reflection or choice, forcing the perceptually encountered scene to appear determinately one way rather than another[7]. In PP, this means that the generative model includes an 'expectation' that the visual scene will be unitary-coherent. Such an expectation may be realized in many different ways, and might itself be the result of early learning or innate pre-disposition. Brains thus constituted still rely on multi-level probabilistic calculations, but are constrained to select, moment-by-moment, a single best-fit unitary model poised for the control of action and choice.

4. An Illustration: Binocular Rivalry

As an illustration, consider the case of 'binocular rivalry'. Binocular rivalry experiments[8] originated with Charles Wheatstone (1838), who invented the 'stereoscope' – a device that used mirrors to present different images to each eye. In contemporary work, the same effect is usually achieved (see e.g. Leopold and Logothetis, 1999; Alais and Blake, 2005) using the kinds of anaglyph image viewing spectacles familiar from early 3D movies and comics. These have differently colored lenses for each eye (e.g. red one eye, cyan the other). This allows the presentation of two different images (one in each colour) at roughly the same on-screen location, while ensuring that each eye receives information about just one of the images. In the (non-rivalrous) case of 3D viewing, the two images are nearly identical, but reflect slightly different perspectives with offsets that create apparent differences in depth. Here, the brain is easily able to merge the two bodies of visual evidence in a way that yields the familiar experience of a single scene with objects at varying depths.

Binocular rivalry occurs when the images presented to each eye depict different scenes altogether. A much-used preparation here is the house/face stimulus, many examples of which can be found online. Here, the anaglyph image combines the broad outline of a house rendered in one colour with that of a face rendered in the other. Exposed to such a stimulus, subjects do not experience (as in the 3D case) a single scene with objects arranged at various depths. Nor do they experience a stable superposition of (e.g.) house and face. Instead, they report an alternating pattern in which first one scene dominates awareness, and then the other, and then the first again. These face/house transitions are not always sharp, and may involve a gradual 'breaking through' of elements of the other image, before it dominates the previous one, after which the whole cycle repeats ((see e.g. Lee at al, 2005). But crucially, the subjective experience is one in which one coherent scene repeatedly gives way to another.

This intriguing pattern within subjective experience falls neatly into place given the predictive processing framework and the considerations concerning action-selection mentioned above. For as Hohwy et al (2008) point out, the brain here commands two perfectly good models (house/face) each of which is able to predict a large amount of the impinging sensory signal. But neither is able to predict the full signal, since that signal – thanks to the ecologically peculiar circumstances involving differently colored lenses and an anaglyph image – carries roughly equal amounts of information suggesting two normally quite incompatible real-world scenes. This delivers what they describe as an 'epistemological' rationale for the constant switching between house and face percepts.

Thus suppose, for whatever reason, that the brain's first attempt to recruit a hypothesis that minimizes prediction errors with respect to the visual inputs discovers that 'house-at-location X' is effective. This 'best-guess' accommodates a large amount of the present signal (everything delivered via one anaglyph channel). But now, prediction error remains for all the information arriving via the other channel. To accommodate this, a new hypothesis must emerge. Fortunately, there is a very good one readily available. For the 'face-at-location-X' hypothesis minimizes all the remaining prediction error. It now dominates, and the 'face' percept enters subjective experience. But now, prediction error accrues for all the information (previously nicely accommodated) entering via the other anaglyph channel. To minimize this, another new hypothesis must be found. Fortunately, there is one readily available – the 'house' hypothesis!

But why, under such circumstances, do we not simply experience a combined or interwoven image: a kind of house/face mashup for example? Although such partially combined percepts do occur, and may persist for brief periods of time, they are never complete (bits of each stimulus are missing) or stable. Such mash-ups do not constitute a viable hypothesis given our more general knowledge about the visual world. For it is part of that general knowledge that, for example, houses and faces do not occupy the same place, at the same scale, at the same time. This kind of general knowledge may itself be treated as a systemic prior, albeit one pitched at a relatively high degree of abstraction (such priors are sometimes referred to as "hyperpriors"). In the case at hand, what is thereby captured is the fact that "the prior probability of both a house and face being co-localized in time and space is extremely small" (Hohwy et al (2008) p. 691). This, indeed, may be the deep explanation of the existence of competition between the higher-level hypotheses in the first place - these hypotheses must compete because the system has learned that "only one object can exist in the same place at the same time" (Hohwy et al (2008) p. 691). The constant switching that characterizes our subjective experience in binocular rivalry cases is thus explained. The switching is the inevitable result of the probabilistic prediction error minimizing regime *as constrained* by the hyperprior that the world in which we live and act in is one in which unitariness and coherence are the default.

5. Intentions, Predictions, and Imagination

What about the parallel puzzle concerning human intentions? Are these, as Holton (2016) suggests, also 'all-or-nothing' states, and if so, does that pose a harder puzzle for the predictive processing story?

Holton's suggestion is that an agent, when she intends to do X, intends just that: to do X. She does not intend to do X with 60% probability and Y (or just 'something else') with 40% probability. Subjectively speaking, this case seems far less clear-cut. Some folk (myself included), may quite explicitly intend to do something with some associated probability. For example, I may say, if asked, that there is a 60% probability that I'll come to a party. And while this may sometimes reflect uncertainties about other commitments, it may just as easily reflect uncertainties about my own future trajectories and desires. But for the sake of argument, let's assume that very often, at least, our own intentions strike us as 'all-or-nothing'. I say to myself "I am going to the party" and I intend simply that - with whatever implied 'ceteris paribus' caveats common-sense dictates. Is this kind of personal-level state consistent with its putative thoroughly probabilistic sub-personal underpinnings?

I think it is. The key point to notice is that intentions and desires, as implemented within the predictive processing schema, involve predictions and imaginative explorations of our own future actions (Pezzulo et al. 2015). That means that the very same pattern of constraints applies. I cannot actually just go 60% to the party, but must end up either going or not going. So my predictions of my own future actions and my imaginative explorations of their consequences, ought to be constrained by this (as by any other) real-world fact.

The realm of mental exploration thus inherits the webs of constraint and possibility proper to their real-world counterparts. In order to make up my mind about the party, I must imaginatively explore possible futures that fall into two very definite camps – one camp features going to the party, and one does not. Once I have decided which (given any other constraints) appeals most, I will end up predicting that I will take one or the other path. The 'all-or-nothing' nature of many human intentions and desires may thus follow directly from the all-or-nothing nature of human action itself.

6. Experiencing Sensory Uncertainty

A residual puzzle concerns the experience of perceptual uncertainty. For we do sometimes experience a lack of certainty about our own intentions, or about what is in the visual scene. Does the picture of visual experience (and perceptual experience more generally) as unitary-coherent render such experiences themselves problematic?

A full treatment of this issue is beyond the scope of this paper, but we may sketch the general shape of a response. Agentive choice and reason needs to be informed by just the right amount of 'personal-level', agent-accessible information – not too much, and not too little. Thus consider the agentive experience of surprise at a certain sight or sound. This can seem puzzling, since to experience the sight or sound *at all* requires (if these stories are on track) the brain to deliver the percept that is least 'surprising' insofar as it best minimizes prediction error given the priors and the current waves of sensory evidence. Plausibly, however, the agentive experience of surprise marks something subtly different: the distance between systemic prior expectations and the least-surprising posterior consequent upon receiving sensory evidence. In this way, the agent is simultaneously apprised both of the current overall (least surprising) best-guess about the world, *and* the fact that this best-guess represents a state of affairs that was previously considered improbable. It is plausible to suggest that the agentive experience of surprise tracks this prior improbability, delivering (potentially important) information concerning the large divergence between prior and posterior.

Similarly, the experience of sensory uncertainty may function to maintain an agent's awareness of her own (sub-personally computed) confidence in current attempts to infer a course of action, or the contents of the distal scene[9]. For example, she may be unsure if the shape seen in the field at dusk is a fox or a dog[10]. Notice (and see Morrison (2016) (2017), and Denison (2017) for further discussion) that this is unlike the case of binocular rivalry. In the case of viewing the shape in the field, we do not normally alternate between seeing a clear and distinct dog-image and seeing a clear and distinct fox-image. Rather, we experience a single, persisting, determinate shape (upon which we might act, by throwing a ball) which is actually quite well accounted for by either one of the two top-level hypotheses. That hard-to-categorize shape is itself the determinate, unitary, best-guess poised for the control of action and choice – even if the action is to move closer to get better information. In binocular rivalry, by contrast, the system commands two equally good best-guesses, each resulting in a clear and distinct hypothesis, but each based on a different body of simultaneously-presented sensory evidence.

It is worth noting that, given that a key role of the Bayesian processing is to drive action, many of those actions will themselves be epistemically motivated. They will be actions whose role is to gather information so as to reduce underlying uncertainty. Because PP stories (like all Bayesian accounts) 'inherently represent the degree of uncertainty' (Kruschke (2008) p.211) the probabilistic processing regime remains positioned to drive behaviors – such as patterns of saccade around the scene- that seek to reduce uncertainty, more clearly favouring one take on the world over another. This means that the active being, though constantly presented with a unitary-coherent scene, is nonetheless constantly engaging that scene in ways that reflect the underlying swathe of probabilistic information. Agents like that get to have their cake and eat it – they encounter a unitary-coherent world as an arena for action and choice, while constantly checking that unitary-coherent take by exploiting, in action, the uncertainties that characterize the underlying Bayesian regime[11].

Finally, an interesting issue that here arises concerns the phenomenal character of peripheral vision[12]. It is sometimes suggested that peripheral visual content - unlike foveal experience - is in some sense 'statistical' in character, or phenomenally indeterminate (see e.g. Seth (2014) pp 113-114, Seth (2017, especially section 2.4), Cohen and Dennett (2011), Kouider et al (2010), Lettvin (1976), Madary (2013)). Such a picture potentially fits with the considerations concerning epistemic action, where the role of repeated foveations is to reduce high-level uncertainty by harvesting low-level information through repeated environmental probes[13]. In a multi-level, multi-area processing regime, support for such 'mixed' phenomenologies might be provided by the deployment of hierarchical generative models whose upper reaches trade in competing discrete state spaces while the lower ones track continuous variations in e.g. luminance, speed, motion, contrast brightness etc. Peripheral vision is well-suited to the task of identifying useful cues using some of these features. Such an architecture amounts, however, to a bag of models that cannot itself have a single peak as there is no common metric space in which to compare them. Nonetheless, the higher levels (negotiating discrete state spaces) would be constrained to interact so as to deliver a unitary-coherent perceptual world fit for action and choice. Conflicting intuitions concerning the probabilistic or non-probabilistic nature of daily visual experience might then be explained by the operation of just such a heterogeneous hierarchical organization, combining continuous and discrete models in ways constrained to serve the ongoing selection of both practical and epistemic actions (for some preliminary discussion, see Friston et al (2017)).

6. Conclusions: Bridging the Gap

There is an apparent tension between the emerging picture of human brains as approximating probabilistic Bayesian inference and some mundane facts about experience itself. If much of the processing underlying experience trades (explicitly or implicitly) in probabilities, why is that character so often hidden from agentive sight? The answer, I have argued, is that perception is always and everywhere in the service of action, and that action (in a world like ours) means choosing between competing but unitary-coherent ways the world might be. Agentive choice and action is thus best served by strongly predicting a unitary-coherent scene, while constantly probing, testing, and revising our percepts in ways driven by self-estimated uncertainty.

Does this mean that the probabilistic picture is of only limited use in *explaining* the personal-level facts? Quite the opposite seems to be true. For it is the deep probabilistic underpinnings of unitary-coherent human experience that explain otherwise puzzling phenomena arising within subjective experience itself, such as the alternating percepts experienced during binocular rivalry[14], and that enables us to engage in constant epistemic actions whose role is to check and improve the unitary-coherent grip itself.

* This paper was written thanks to support from ERC Advanced Grant XSPECT - DLV-692739. Special thanks to Anil Seth and Karl Friston, and to two anonymous referees whose careful critiques helped clarify the focus and greatly improve the argument.

References

Alais, D & Blake, R. 2005. *Binocular rivalry*. MIT Press, Cambridge, MA

Churchland, P.S., Ramachandran, V., & Sejnowski, T. (1994). A Critique of Pure Vision. In C. Koch & J. Davis (Eds.), *Large-Scale Neuronal Theories of the Brain* . Cambridge, MA: MIT Press.23-61

Clark, A. 2013. Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science *Behavioral and Brain Sciences* 36: 3: p. 181-204

Clark, A. 2016. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind* (Oxford University Press, NY)

Cohen, M. A., & Dennett, D. C. 2011. Consciousness cannot be separated from function. *Trends Cogn Sci*, 15(8), 358–364.

Dennett DC. 1969. *Content and consciousness*. Routledge and Kegan Paul, London.

Denison, R. N. 2017. Precision, Not Confidence, Describes the Uncertainty of Perceptual Experience: Comment on John Morrison's "Perceptual Confidence".

Analytic Philosophy, 58(1), 58-70.

Friston, K. 2010. The free-energy principle: a unified brain theory? *Nature Reviews. Neuroscience*, *11*(2), 127–38

Friston, K 2011 What is optimal about motor control? *Neuron* 72: 488-498

Friston, K. J., Parr, T., & de Vries, B. 2017. The graphical brain: belief propagation and active inference. *Network Neuroscience*, 1–78.

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. 2015. Active inference and epistemic value. *Cognitive Neuroscience*, 6(4), 187–214.

Friston K, Samothrakis S, and Montague R 2011 Active inference and agency: optimal control without cost functions *Biol Cybern.* 106 (8-9) 523-541

Hohwy, J. 2013. *The Predictive Mind* (Oxford University press, NY)

Hohwy, J., Roepstorff, A., Friston, K. 2008. Predictive coding explains binocular rivalry: an epistemological review. *Cognition* 108 (3): 687-701.

Holton, R. 2016. Review of Surfing Uncertainty: Prediction, Action, and the Embodied Mind, *Times Literary Supplement* October 7, 10-11

Klein, C. (In Press). What do predictive coders want? *Synthese.*

Kouider, S., de Gardelle, V., Sackur, J., & Dupoux, E. 2010. How rich is consciousness? The partial awareness hypothesis. *Trends in Cognitive Science,* 14 (7), 301–307.

Kruschke, J. K. (2008). Bayesian approaches to associative learning: from passive to active learning. *Learning & Behavior* 36(3), 210–226.

Knill, D., and Pouget, A. 2004. The Bayesian brain: the role of uncertainty in neural coding and computation *Trends in Neuroscience*. 27(12):712-9

Le Clerc, S. 1712. *Système de la vision*. Paris: Delaulne.

Lee, S.H., Blake,R., & Heeger,D.J. 2005.Traveling waves of activity in primary visual cortex during binocular rivalry. *Nature Neuroscience* 8, 2.

Leopold D.A., Logothetis N.K. 1999. Multistable phenomena: changing views in perception. *Trends Cogn Sci* 3: 254-264.

Lettvin, J. Y. 1976. On seeing sidelong. *The Sciences,* 16, 10–20.

Lu, Stafford, and Fox. 2016. Maximum Saliency Bias in Binocular Fusion *Connection Science* 28:3: 258-269

Lupyan, G. 2015. Cognitive Penetrability of Perception in the Age of Prediction: Predictive Systems are Penetrable Systems. *Review of Philosophy and Psychology*, 6 (4):547-569

Madary, M. 2012. How would the world look if it looked as if it were encoded as an intertwined set of probability density distributions?. *Frontiers in Psychology*, 3, 419.

Marr, D. 1982. *Vision: A computational approach*. Freeman

Michael, E., de Gardelle, V., & Summerfield, C. (2014). Priming by the variability of visual information. *Proceedings of the National Academy of Sciences*, 111(21), 7873–7878.

Morrison, J. 2016. Perceptual Confidence. *Analytic Philosophy*, 57(1), 15-48.

Morrison, J. 2017. Perceptual Confidence and Categorization. *Analytic Philosophy*, 58(1), 71-85.

Pezzulo, G., Rigoli, F., & Friston, K. 2015. Active Inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, *134*, 17–35

Porta, J. B. 1593. *De refractione. Optices parte. Libri novem*. Naples: Salviani.

Seth, A. K. 2013. Interoceptive Inference, Emotion, and the Embodied Self. *Trends in Cognitive Sciences* 17, no. 11: 565–73.

Seth, A. K. (2014). A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive Neuroscience*, *5*(2), 97–118. http://doi.org/10.1080/17588928.2013.877880

Seth, A. (2017, August 7). From unconscious inference to the beholder's share: predictive perception and human experience. Retrieved from osf.io/preprints/psyarxiv/zvbkp

Shipp S., Adams R. A., Friston K. J. 2013. Reflections on agranular architecture: predictive coding in the motor cortex. *Trends in Neuroscience* 36, 706–716

Wiese, W. (2016). What are the contents of representations in predictive processing? *Phenomenology and the Cognitive Sciences*, 1-22.

---

[1] For reviews and introductions, see Clark 2013, 2016, Friston 2010, Hohwy 2013.

[2] Thanks to an anonymous referee for suggesting I clarify the difference that various ways of approximating Bayesian inference make to this picture, and to the puzzle I later target.

[3] 'Bayesian' because these systems combine prior knowledge with new evidence in ways that approximate Bayesian inference.

[4] Thanks to an anonymous referee for pursuing this issue.

[5] Thanks to an anonymous referee for stressing the importance of showing why this simple and popular move does not resolve the main worry.

[6] For proximal action control, this works roughly as follows. A predicted action is unpacked into a sequence of predicted proprioceptive states. Since these states are not actual, prediction errors emerge. Those errors are then quashed by bringing the action about. Proprioceptive predictions thus play the role of motor commands, and they act as a kind of 'self-fulfilling prophecy' that entrains the action - see Shipp et al 2013, and for a non-technical treatment, see Clark 2016: chapter 4. Longer-term control of action (for example,

by standing intentions) is treated in much the same way except that here, the entraining predictions concern patterns in the unfolding of events at a much coarser scale. A satisfying account of this challenging proposal is beyond the scope of this short treatment -see Pezzulo et al. 2015.

[7] Could we not perhaps preserve the puzzle by asking why we do not wait until an action is needed before enforcing a unitary-coherent take. This is a deep and interesting question. The answer, I suspect, is that action is not so much an occasional response to an input as an ongoing art of an integrated perception/action process - a constantly rolling cycle in which perception and action are co-determined and co-determining (see Churchland et al (1994)) with neither one waiting on the other to finish its work before making its own contributions. For more on how PP fits with this kind of vision, see Clark (2016) chapter 8 (Thanks to an anonymous reviewer for raising this issue).

[8] The phenomenon itself was noticed, in less controlled ways, by many others including Porta (1593) and Le Clerk (1712).

[9] This is reflected in the variable weighting of prediction error signals – see Clark 2016: chapter 2.

[10] Similarly, someone may feel unsure about whether or not they intend to go to the party, where that uncertainty reflects systemic lack of confidence in the inference whose conclusion is to go, and in the opposing inference too.

[11] For a compelling picture of the way Bayesian approaches drive behaviors that actively interrogate the environment so as to reduce uncertainty, see Kruschke (2008), and for the PP spin on that story, see Friston et al (2015).

[12] Thanks to Anil Seth for discussion of this issue, and to Karl Friston for pointing me to the work on continuous and discrete state spaces.

[13] See also Michael et al (2014), who show priming effects that reflect rapid perceptual estimations of variance itself, providing contextual clues that inform ongoing perception.

[14] For many further examples of the power of the probabilistic story to illuminate personal-level experience, see e.g. Hohwy 2013, Clark 2013, 2016, Seth 2013, Lupyan 2015. Our personal-level unitary-coherent percepts, as large bodies of work in psychophysics (see Knill and Pouget 2004) further attest, are themselves best explained by a picture of the brain as a near-optimal Bayesian integrator of multiple streams of probabilistic sensory information.